CubeView: A System for Traffic Data Visualization

S. Shekhar, C. T. Lu, R. Liu, C. Zhou
Computer Science Department, University of Minnesota
200 Union Street SE, Minneapolis, MN-55455
[*shekhar*, *ctlu*, *rliu*, *czhou*]@cs.umn.edu TEL:(612)624-8307 FAX:(612)625-0572
http://www.cs.umn.edu/Research/shashi-group

April 11, 2002

Abstract

Due to the huge amounts of data being stored in traffic and transportation databases as well as many other information repositories, traffic data visualization is receiving substantial interest from both academia and industry. The Minneapolis-St. Paul(Twin-Cities) traffic archival stores sensor network measurements collected from the freeway system in the Twin-Cities metropolitan area. In this paper, we present a web-based visualization package for observing rapid summarization of major trends. In the underlying database, we modeled the traffic data as a multi-dimensional data warehouse to facilitate the on-line query processing used in the visualization software. We also discuss some research issues in mining traffic and transportation data.

Keywords: Data Visualization, Traffic Data, Data Mining

1 Introduction

Transportation agencies and companies are currently collecting huge amount of data concerning their operations and systems. Until recently it has been difficult to discover trends, find patterns, or locate anormalities within these large data sets. Data visualization is expected to play an important role in extracting such useful information and thus provide a means for agencies and companies to make more effective decisions.

Data visualization techniques are particularly important for Intelligent Transportation Systems(ITSs), which deal with the sensors, instrumentation, communication and control of the transportation system. Current ITS data is primarily used for real-time decision making. It has not been extensively and systemically applied for long-term data analysis and decision-making.

In this paper, we formulate a general framework for visualizing transportation data. We present a web-based visualization package for observing rapid summarization of major traffic trends. In the underlying database, we model the traffic data as a multi-dimensional data warehouse to facilitate the on-line query processing used in the visualization software. We also discuss some research issues in mining traffic and transportation data.

The rest of the paper is organized as follows. Section 2 introduces the traffic data archival of one urban region in the United States. Section 3 presents our *CubeView* visualization system. Section 4 demonstrates the design of our software system. Section 5 illustrates how a variety of users can benefit from this visualization tool. Finally, we summarize our work in Section 6.

2 Twin-Cities Traffic Data Archival

In 1995, the University of Minnesota and the Traffic Management Center(TMC) Freeway Operations group started the development of a database to archive sensor network measurements from the freeway system in the Twin Cities of Minneapolis and St. Paul. The sensor network includes about nine hundred stations, each of which contains one to four loop detectors, depending on the number of lanes. Sensors embedded in the freeways and interstate monitor the occupancy and volume of traffic on the road. At regular intervals, this information is sent to the Traffic Management Center for operational purposes, e.g., ramp meter control, as well as research on traffic modeling and experiments. Figure 1 shows a map of the stations on highways within the Twin-Cities metropolitan area, where each small polygon represents one station. The interstate freeways include I-35W, I35E, I-94, I-394, I-494, and I-694. The state trunk highways include TH-100, TH-169, TH-212, TH-252, TH-5, TH-55, TH-62, TH-65, and TH-77. I-494 and I-694 together form a ring around the Twin-Cities. I-94 passes from East to North-West and I-35W and I-35E run in a South-North direction. Downtown Minneapolis is located at the intersection of I-94, I-394, and I-35W, and downtown St. Paul is located at the intersection of I-35E and I-94.

With the huge amount of traffic data stored in its traffic database, the Minnesota Department of Transportation(MNDOT) needed a high-performance traffic data visualization system that extracts patterns and rules from the historical data to support decision making.

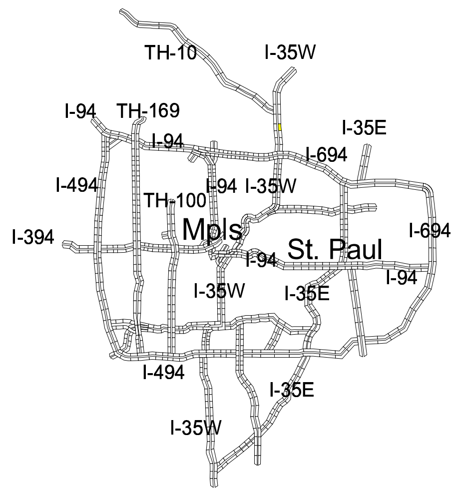


Figure 1: Detector Map of Stations on Twin-cities Highways

3 CubeView

Based on the needs from MNDOT and requirement analysis, we construct the *CubeView* visualization sytem. *CubeView* presents information in various formats to assist transportation managers, traffic engineers, travellers and commuters, and researchers and planners to observe and analyze traffic trends.

In general, the subjects of analysis in a multidimensional data model are a set of numeric measures. Each of the numeric measures is determined by a set of dimensions. In a traffic data warehouse, for example, the measures are volume and occupancy, and the dimensions are *time* and *space*. Dimensions are hierarchical by nature. For example, the *time* dimensions can be grouped into "Week","Month","Season", or "Year", which form a lattice structure indicating a partial order for the dimension. Similarly, the *Space* dimensions can be grouped into "Station","County","Freeway", or "Region". Given the dimensions and hierarchy, the measures can be aggregated in different ways. For example, for a particular highway and a chosen month, the weekly traffic volumes can be analyzed. The SQL aggregate functions and the group-by operator only produce one out of all possible aggregates at a time. A data cube is an aggregate operator which computes all possible aggregates in one shot. The CUBE operator generalizes the histogram, cross-tabulation, roll-up, drill-down, and sub-total constructs. It is the N-dimensional generalization of simple aggregate functions.

For the Twin Cities traffic data, we have a cube view as a specialization of general data cube operator as in Figure 2. In this figure, T_{TD} represents the time of a day, T_{DW} represents the day of a week, T_{MY} represents the month of a year and S represents the station. Each node is a data cube operation and a view of the data. For example, S represents the traffic volume of each station of all the time. ST_{TD} represents daily traffic volume of each station. $T_{TD}T_{DW}S$ represents traffic volume on each station of different time of a day, which is generated as a video in *CubeView* system.

The basic concept behind *CubeView* is data cube. Because of the nature of data cube, *CubeView* can analyze any traffic data. The software only requires space and time for each measure, like volume, occupancy and speed. Speed is derived from volume and occupancy. These requirements are very simple and most highway monitoring systems should be able to satisfy them.

We use Figure 3 to illustrate the data flows and the required modules of the *CubeView* system. The basic map and raw data are cleaned, transformed, and loaded into the data warehouse module, which provides the multidimensional views and the Online Analytical Processing(OLAP) operations for data visualization as well as a variety of data mining analysis tools, e.g, classification, clustering, outlier detection. The discovered patterns or rules are then visually displayed as maps or charts for further interpretation.

The *CubeView* system is accessed through a Web browser and provides a user-friendly interface. Different kinds of users can benefit from the system with their particular needs. Transportation managers can assess the effect of ramp meter control using the traffic comparison components; traffic engineerer can analyze the abnormal traffic flow through the outlier analysis tools; travelers can observe the historical traffic flow for a specific event and avoid congestion road segments.

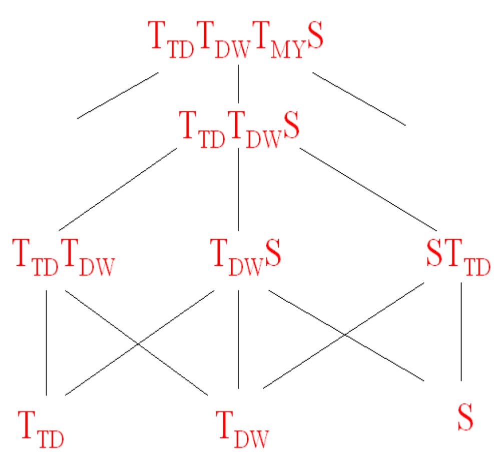


Figure 2: Cube View For Traffic Data

4 Software Architecture

The design of the software is following the scalable 3-tier architecture:

- Graphic User Interface (GUI): The GUI draws a highway map using the geographic coordination information of each station. It accepts queries from users and sends queries to middle tier, which further requests the traffic data from the database tier. The GUI can display traffic video, detect outlier stations, and show highway volume maps for a user-specified time, date, and highway stations.

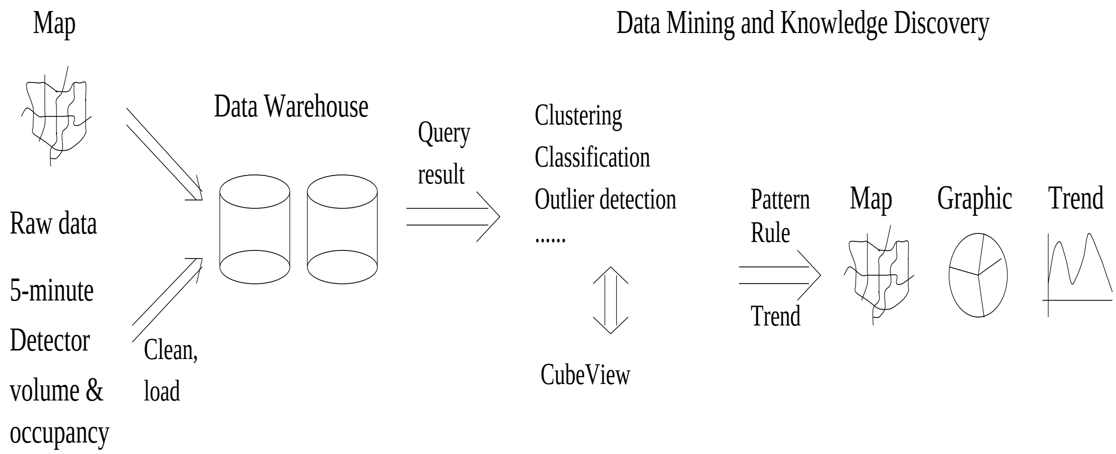


Figure 3: Data-flow and Main Modules in Our System

- Data Access Middle Tier: The middle tier accepts requests from GUI and retrieves data from database tier. Performance optimization is conducted on this tier to optimize high performance of the system.
- Database Tier: The database tier stores the 5-minute interval road detection data and the descriptions of each station. Different indexing and design strategies are used in database tier for high performance data retrieval.

Figure 4 shows the architecture of our software system. Users can access our system through browser and Internet. The web server and database server are located at University of Minnesota and are connected through fast ethernet.

Using Web browser as user interface allows users to access our software worldwide. The user interface also provides modern control components to maximize the interactivity using popular Java techniques. Having only Internet connection and a browser, traffic masters can analyze traffic data anywhere.

In *CubeView*, raw data collected by loop detectors are stored in binary format. The binary data are converted into text data using transformation programs and later inserted into database servers.

5 Users of Transportation Traffic Visual Tools

Transportation Managers

Besides to be able to access real time traffic data, transportation managers also need to be able to analyze traffic data spatially and temporally. Traffic patterns need to be discovered and trends need to be concluded. Using *CubeView*, analysis and research tasks can be easily and reapidly accomplished. Figure 5 shows an example of a comparison video snapshot. In this figure, two traffic video snapshots of two different days are displayed. Transportation managers can watch the videos simultaneously and detect abnormal situations right away. A decision could be made beased on the patterns and resources can be allocated to improve the traffic in a

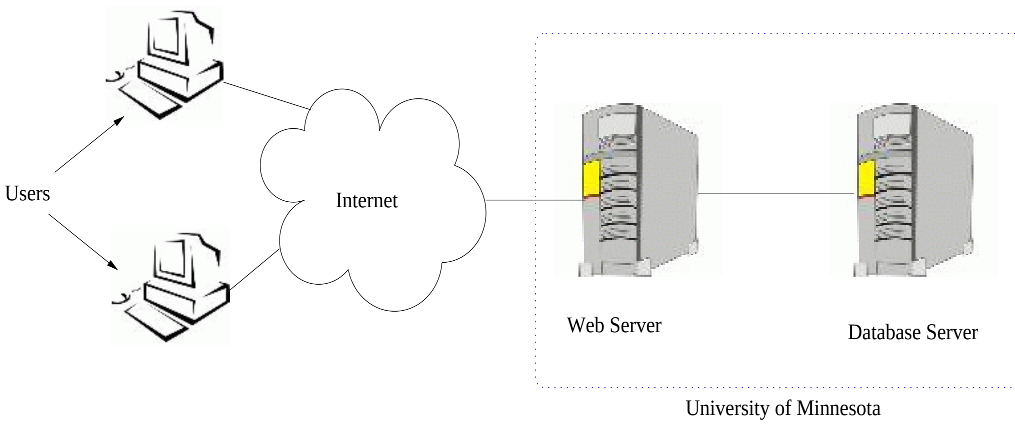


Figure 4: Software Architecture

short turn around time. The video operation represents the $T_{TD}T_{DW}S$ node in the *CubeView* in Figure 2.

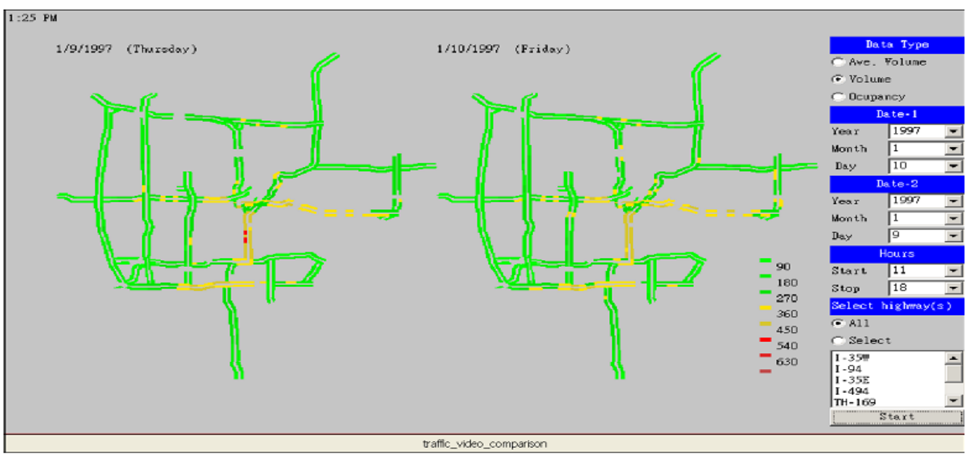
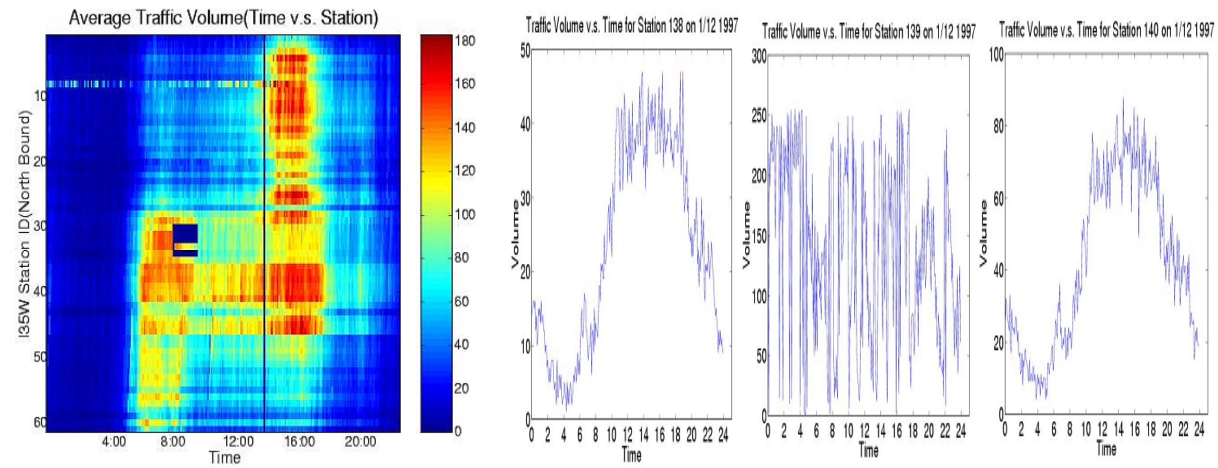


Figure 5: A Comparison of Traffic Video Snapshot

Traffic Engineers

Traffic engineers are responsible for a wide variety of traffic issues, such as collecting and evaluating data related to traffic flow, traffic speeds, circulation patterns, roadway capacity, sight visibility and traffic accidents; evaluating the need for traffic signals, etc. *CubeView* provides a set of tools to accomplish those goals. For example, Figure 6 (a) and (b) display the daily traffic volume on 3 different stations. From Figure 6 (a), abnormal traffic spots(dark blocks) can be easily detected. These abnormal traffic pattern can be viewed in detail in charts in Figure 6 (b). This operation represents the ST_{TD} node in the *CubeView* in Figure 2.



(a) Outlier Volume Map

(b) Outlier Station Detectors

Figure 6: Outlies in Different Picture

Travelers and Commuters

CubeView can help commuters select their commuting routes by comparing the volume of different highway at the same time. The results are rendered using charts, which are very user-friendly. Figure 7 shows that 35W/280 has less traffic compared with others on everyday. It might be the ideal comuting route. Right portion of Figure 8 is a view of traffic occupancy during the day for each station. Congestion can be detcted visually. By clicking on the congestion area(dark blocks), the corresponding congested highways are highllighted on the map in the left portion.

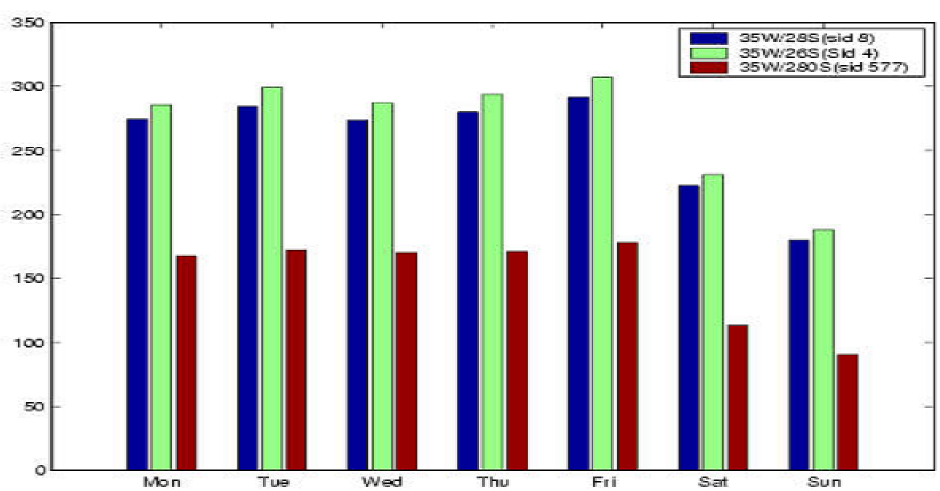


Figure 7: Traffic Volume of Different Highway on Each Weekday

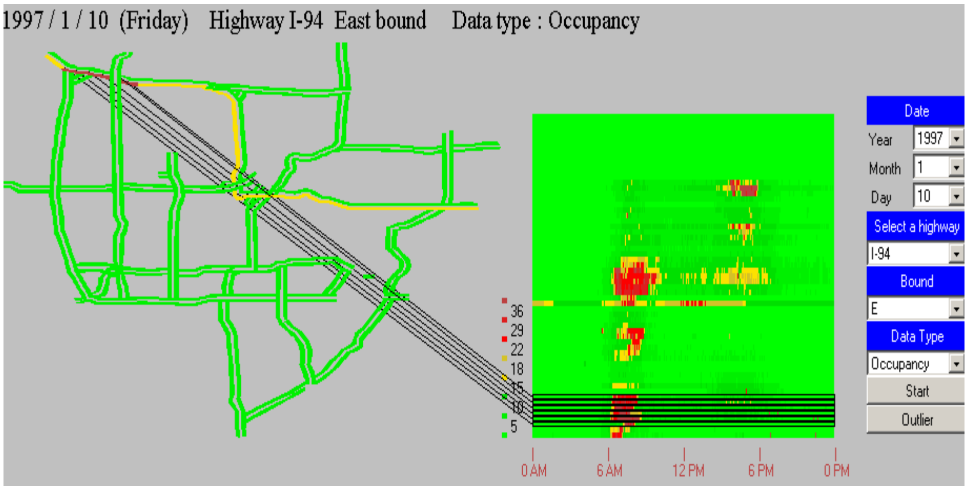


Figure 8: Congestion Detection

Researchers and Planners

Spending millions and millions of dollars to build more and more roads is simply not an acceptable solution for improving traffic conditions. One solution traffic planners can look to is information technology, which can help them discover traffic patterns and improve traffic condition. Figure 9 classifies traffic using clustering classification. From the classification, we can tell that the highways in downtown area have a higher volume. In some cases, planners also need to build models to mimic the traffic resulted from different impacts such as new development, traffic signal timing, weather condition, etc. Integrated with classical data mining techqiues, *CubeView* gives both researchers and planners a new powerful tool to accomplish these research and planning.

6 Conclusions

Every day, a huge amount of traffic data is collected in every state in the United States and stored in storage such as tape, disk, and CD. It is known that statistics from this data has great potential to improve the traffic performance. However, it is tedious, if not often impossible, for normal users to understand the meaning of these statistics. Even traffic professionals find it quite difficult to capture useful information from huge collections of statistic data. Most travellers do not have the time or energy to determine the location of congestion by interpolating data on volume and occupancy per five minutes. In fact, travellers generally just want to know traffic conditions on a particular freeway within a certain range and time just by glancing at a picture. Information technology can provide such "simple" but useful information quite easily.

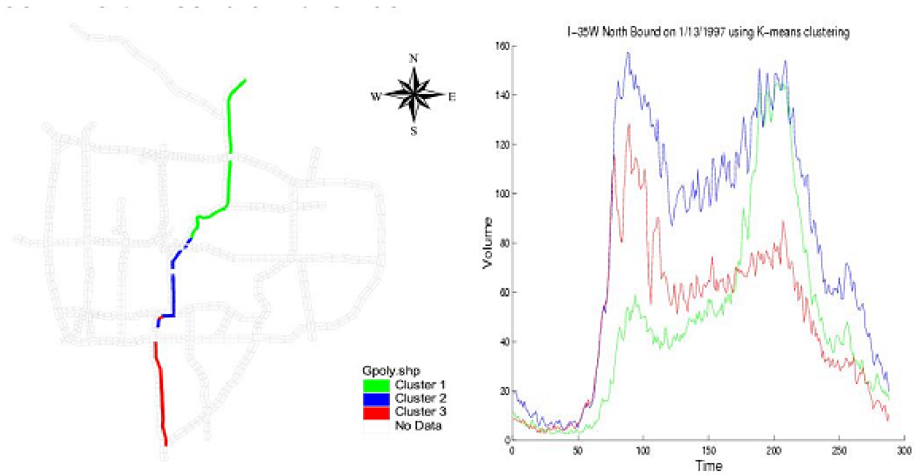


Figure 9: Traffic Classification

Internet based traffic visualization tools provide an easy accessible approach for reducing complex and tedious statistical data to simple but powerful information that can benefit non-professional and professional users alike.

7 Acknowledgments

We are particularly grateful to our Spatial Database Group members, Weili Wu, Hui Xiong, Xiaobin Ma, and Pusheng Zhang, for their helpful comments and valuable discussions.

This work is supported in part by a USDOT grant on high performance spatial visualization of traffic data, and is sponsored in part by the Army High Performance Computing Research Center under the auspices of the Department of the Army, Army Research Laboratory cooperative agreement number DAAH04-95-2-0003/contract number DAAH04-95-C-0008, the content of which does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. This work was also supported in part by NSF grant #9631539.

References

[1] In *URL: http://www.census.gov.*

[2] S. Chaudhuri and U. Dayal. An overview of data warehousing and olap technology. In *Proc. VLDB Conference*, page 205, 1996.

[3] S. Chaudhuri and U. Dayal. An overview of data warehousing and olap technology. (1):65–74, March 1997.

[4] E.F. Codd. Twelve rules for on-line analytic processing. In *Computerworld, April 13 1995.*

[5] E.F. Codd, S.B. Codd, and C.T. Salley. Providing olap(on-line analytical processing) to user-analysts: An it mandate. In *Arbor Software Corporation. Avaliable at http://www.arborsoft.com/essbase/wht_ppr/coddToc.html*, 1993.

[6] ESRI. Spatial data warehousing. http://www.esriau.com.au/wp.htm.

[7] P. Ferguson. Census 2000 behinds the scenes. In *Intelligent Enterprise*, October 1999.

[8] J. Gray, A. Bosworth, A. Layman, and H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-total. In *Proceedings of the Twelfth IEEE International Conference on Data Engineering*, pages 152–159, 1995.

[9] J. Han, N. Stefanovic, and K. Koperski. Selective materialization: An efficient method for spatial data cube construction. In *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'98)*, pages 144–158, 1998.

[10] W.H. Inmon. *Building the Data Warehouse*. New York, NY: John Wiley & Sons, 1993.

[11] W.H. Inmon and R.D. Hackathorn. *Using the Data Warehouse*. New York, NY: John Wiley & Sons, 1994.

[12] W.H. Inmon, J.D. Welch, and K.L. Glassey. *Managing the Data Warehouse*. New York, NY: John Wiley & Sons, 1997.

[13] J. Widom. Research Problems in Data Warehousing. In *Proc. of 4th Int'l Conf. on Information and Knowledge Management(CIKM), Nov. 1995.*

[14] R. Kimball, L. Reeves, M. Ross, and W. Thornthwaite. *The Data warehouse Lifecycle Toolkit*. Wiley, 1998.

[15] MICROSOFT. Terraserver: A spatial data warehouse. http://www.microsoft.com.

[16] I.S. Mumick, D. Quass, and B.S. Mumick. Maintenance of data cubes and summary tables in a warehouse. In *SIGMOD*, pages 100–111, 1997.

[17] S. Shekhar, S. Chawla, S. Ravada, A. Fetterer, X. Liu, and C.T. Lu. Spatial databases: Accomplishments and research needs. *IEEE Transactions on Knowledge and Data Engineering(TKDE)*, 11(1):45–55, 1999.

[18] M. Stonebraker, J. Frew, and J. Dozier. The sequoia 2000 project. In *Proceedings of the Third International Symposium on Large Spatial Databases*, 1993.

[19] USGS. National satellite land remote sensing data archive. In *http://edc.usgs.gov/programs/nslrsda/overview.html*.